An Invariance Principle For Polytopes

Prahladh Harsha Adam Klivans Raghu Meka

The University of Texas at Austin
{prahladh,klivans,raghu}@cs.utexas.edu

Abstract

Let X be randomly chosen from $\{-1,1\}^n$, and let Y be randomly chosen from the standard spherical Gaussian on $\mathbb{R}^n$. For any (possibly unbounded) polytope P formed by the intersection of k halfspaces, we prove that

$$|\Pr[X \in P] - \Pr[Y \in P]| \leq \log^{8/5} k \cdot \Delta,$$

where Δ is a parameter that is small for polytopes formed by the intersection of "regular" halfspaces (i.e., halfspaces with low influence). The novelty of our invariance principle is the polylogarithmic dependence on k. Previously, only bounds that were at least linear in k were known.

We give two important applications of our main result:

- A bound of $\log^{O(1)} k \cdot \varepsilon^{1/6}$ on the Boolean noise sensitivity of intersections of k "regular" halfspaces (previous work gave bounds linear in k). This gives a corresponding agnostic learning algorithm for intersections of regular halfspaces.
- A pseudorandom generator (PRG) with seed length $O(\log n \operatorname{poly}(\log k, 1/\delta))$ that δ-fools *all* polytopes with k faces with respect to the Gaussian distribution.

We also obtain PRGs with similar parameters that fool polytopes formed by intersection of regular halfspaces over the hypercube. Using our PRG constructions, we obtain the first deterministic quasi-polynomial time algorithms for approximately counting the number of solutions to a broad class of integer programs, including dense covering problems and contingency tables.

1 Introduction

An important theme in theoretical computer science over the last two decades has been the usefulness of translating a combinatorial problem over a discrete domain (e.g., $\{-1,1\}^n$) to a problem in continuous space. The notion of convex relaxation, for example, is now a standard technique in the design of algorithms for optimization problems. More recently, the study of analytic properties of Boolean functions (e.g., Fourier spectra and sensitivity) has been a fundamental tool for proving results in hardness of approximation [Wol08, O'D08] and learning theory [Man94].

The influential work of Mossel, O'Donnell, and Oleszkiewicz [MOO05] proving the "Majority Is Stablest" conjecture has led to a rich collection of hardness results for constraint satisfaction problems, most notably for the Max-Cut problem. The crux of their work is an invariance principle relating the behavior of low-degree polynomials over the uniform measure on $\{-1,1\}^n$ to their behavior with respect to Gaussians:

Theorem 1.1 (invariance principle [MOO05])**.** *Let P be a multilinear polynomial such that $\|P\| = 1$. Then, for any $t \in \mathbb{R}$,*

$$\left| \Pr_{x \in_u \{-1,1\}^n}[P(x) > t] - \Pr_{x \leftarrow \mathcal{N}^n}[P(x) > t] \right| \leq \tau.$$

Here $\mathcal{N}^n$ is the standard multivariate spherical Gaussian distribution on $\mathbb{R}^n$; the parameter τ depends on the coefficients of P and is small if P is "regular" in the sense that the "influence" of each variable in P is small.

Roughly speaking, the above invariance principle says that the cumulative distribution function (cdf) of a polynomial over $\{-1,1\}^n$ is close to the cdf of a polynomial over $\mathcal{N}^n$ if the coefficients of the polynomial are sufficiently regular. Although developed in the context of hardness of approximation, their invariance principle has also found applications in the theory of social choice [O'D08] and more recently to the construction of pseudorandom generators for polynomial threshold functions [MZ09].

The main result of this paper is an invariance principle for characteristic functions of polytopes. Recall that a polytope $\mathcal{K}$ is a (possibly unbounded) convex set in $\mathbb{R}^n$ formed by the intersection of some finite number of supporting halfspaces. We refer to $\mathcal{K}$ as a k-polytope if it is equal to the intersection of k halfspaces. Our main theorem is as follows:

Theorem 1.2 (invariance principle for polytopes; see Theorem 3.1 for exact statement)**.** *Let $\mathcal{K}$ be a k-polytope. Then,*

$$\left| \Pr_{x \in_u \{-1,1\}^n}[x \in \mathcal{K}] - \Pr_{x \leftarrow \mathcal{N}^n}[x \in \mathcal{K}] \right| \leq \log^{8/5} k \cdot \Delta.$$

The parameter Δ depends on the coefficients of the bounding hyperplanes of $\mathcal{K}$ and is small if these coefficients are sufficiently regular. In particular, if $\mathcal{K}$ equals $\{x \mid W^T x \leq \theta\}$ for $W \in R^{n \times k}$ and $\theta \in R^k$, and each column u of W is ε-regular, i.e., satisfies $\sum_{i=1}^n u_i^4 \leq \varepsilon^2 \|u\|_2^2$, then Δ is less than $\varepsilon^{1/6}$. Note that there is no restriction on the vector θ. Our invariance principle also holds more generally for any product distribution that is hypercontractive and whose first four moments are appropriately bounded.

The novelty of our theorem is the dependence of the error on k. Applying a recent result due to Mossel [Mos08], it is possible to obtain a statement similar to Theorem 1.2 with an error term that has a polynomial dependence on k. Achieving polylogarithmic dependence on k, however, is much harder, and we need to use some nontrivial results from the analysis of convex sets in Gaussian space.

The case $k = 1$, a single halfspace, is equivalent to the classical Berry-Esséen theorem [Fel68], a fundamental theorem from probability and statistics giving a quantitative version of the Central Limit Theorem. We can therefore view our principle as a generalization of the Berry-Esséen theorem for polytopes. Further, understanding the structure of integer points in polytopes (that is, solutions to integer programs) is an important topic in computer science [BV08], optimization [Zie95], and combinatorics [BR07], and we believe our invariance principle will find many applications.

In this paper, we use our invariance principle to derive new results in learning theory and derandomization.

1.1 Results on Noise Sensitivity of Intersections of Halfspaces

Noise sensitivity of Boolean functions introduced in the seminal works of [KKL88], [BKS99] is an important notion in the *analysis* of Boolean functions. Roughly speaking, the noise sensitivity of a Boolean function f measures the probability over a randomly chosen input x that f changes sign if each bit of x is flipped independently with probability δ.

Bounds on the noise sensitivity of Boolean functions have direct applications in hardness of approximation [Hås01, KKMO07], hardness amplification [O'D04], circuit complexity [LMN93], the theory of social choice [Kal05], and quantum complexity [Shi00]. Here, we focus on applications in learning theory, where it is known that bounds on the noise sensitivity of a class of Boolean functions yield learning algorithms that succeed in harsh noise models such as the agnostic model of learning [KKMS08].

A direct application of our invariance principle Theorem 1.2 gives the following new bound on the noise sensitivity of intersections of regular halfspaces:

Theorem 1.3 (learning application: noise sensitivity of intersections of halfspaces)**.** *Let f be computed by the intersection of k, ε-regular halfspaces. Then the Boolean noise sensitivity of f for noise rate ε is at most* $(\log k)^{O(1)} \cdot \varepsilon^{1/6}$.

Applying a result of Kalai et al. [KKMS08] and Klivans et al. [KOS04], the above theorem implies that intersections of k, ε-regular halfspaces are agnostically learnable with respect to the uniform distribution on $\{-1, 1\}^n$ in time $n^{(\log^{O(1)} k)}$ for any constant error parameter. In particular, intersections of $\{-1, 1\}$ halfspaces (oriented majorities) are ε-regular and fall into this class. The previous best algorithm for learning this concept class, even in the easier PAC model, ran in time $n^{O(k^2)}$ ([KOS04, KKMS08]).

The current best bound for the noise sensitivity of intersection of k arbitrary halfspaces is $O(k\sqrt{\varepsilon})$. This bound is obtained by starting with the $\sqrt{\varepsilon}$ noise sensitivity bound for a single halfspace due to Peres [Per04] and applying a union bound over k halfspaces. On the other hand, optimal bounds of $\Theta(\sqrt{\log k}\sqrt{\varepsilon})$ for the related Gaussian noise sensitivity were obtained recently by Klivans et al. [KOS08]. We believe that the right order for Boolean noise sensitivity of intersection of k halfspaces is $\Theta(\sqrt{\log k}\sqrt{\varepsilon})$ as well.

Improving the bounds for Boolean noise sensitivity would be of considerable interest, particularly for the learning theory applications, as learning the class of intersections of halfspaces even with respect to specific distributions such as the uniform distribution over $\{1, -1\}^n$ is an important open problem in learning theory. We feel that our result is an important step towards improving noise sensitivity bounds for intersections of arbitrary (not necessarily regular) halfspaces.

1.2 Results on Pseudorandom Generators for Polytopes

Our invariance principle also yields new results for several problems in derandomization. Recall the following definition of pseudorandom generators (PRGs):

Definition 1.4. *Let μ be a distribution over $\mathbb{R}$. A function $G : \{0,1\}^r \to \{1,-1\}^n$ is said to δ-fool a polytope $\mathcal{K}$ with respect to μ if the following holds.*

$$\left| \Pr_{y \in_u \{0,1\}^r}[G(y) \in \mathcal{K}] - \Pr_{X \leftarrow \mu^n}[X \in \mathcal{K}] \right| \leq \delta.$$

Combining our invariance principle with a PRG similar to a recent construction of Meka and Zuckerman [MZ09], we obtain a black-box algorithm for approximately counting the number of $\{-1,1\}^n$ points in polytopes formed by the intersection of regular halfspaces:

Theorem 1.5 (PRGs for regular polytopes and approximate counting)**.** *For all $\delta \in (0,1)$, there exists an explicit PRG $G : \{0,1\}^r \to \{1,-1\}^n$ with $r = O((\log n \log k)/\varepsilon)$ that δ-fools all polytopes formed by the intersection of k ε-regular halfspaces with respect to all proper and hypercontractive distributions μ for $\varepsilon = \delta^5/(\log^{8.1} k)(\log(1/\delta))$.*

The constants above depend on the hypercontractivity constants of μ. We define proper and hypercontractive distributions in the next section and remark that the uniform distribution over $\{-1,1\}^n$ is an example of such a distribution.

Theorem 1.5 implies quasi-polynomial time, deterministic, approximate counting algorithms for a broad class of integer programs. For example, dense covering programs such as dense set-cover, and $\{0,1\}$-contingency tables correspond to polytopes formed by the intersection of ε-regular halfspaces. For these types of integer programs, we can deterministically approximate, to within an additive error ε, the number of integer solutions in quasi-polynomial time.

As stated, our invariance principle applies to polytopes whose bounding hyperplanes have coefficients that are sufficiently regular. In some cases, however, we can randomly rotate an arbitrary polytope so that all the bounding hyperplanes become regular. As such, after applying a suitable random transformation (which we derandomize), we can build PRGs for *arbitrary* polytopes if the underlying distribution is spherically symmetric (e.g., Gaussian):

Theorem 1.6 (PRGs for Gaussian space)**.** *For a universal constant $c > 0$ and all $\delta > c \log^2 k / n^{1/11}$, there exists an explicit PRG $G_{\mathcal{N}} : \{0,1\}^r \to \mathbb{R}^n$ with $r = O((\log n)(\log^{9.1} k)/\delta^{5.1})$ that δ-fools all k-polytopes with respect to $\mathcal{N}$.*

Additionally, we prove an invariance principle for polytopes with respect to the uniform distribution over the n-dimensional sphere S^{n-1}. This allows us to easily modify our PRG for polytopes in Gaussian space and build PRGs for intersections of spherical caps:

Theorem 1.7 (PRGs for intersections of spherical caps)**.** *For a universal constant $c > 0$ and all $\delta > c \log^2 k / n^{1/11}$, there exists an explicit PRG $G_{sp} : \{0,1\}^r \to S^{n-1}$ with $r = O((\log n)(\log^{9.1} k)/\delta^{5.1})$ that δ-fools all k-polytopes with respect to the uniform distribution over S^{n-1}.*

An immediate consequence of the above PRG construction is a polynomial time derandomization of the Goemans-Williamson approximation algorithm for Max-Cut [GW95] and other similar hyperplane based randomized rounding schemes. Observe that this derandomization is a *black-box* derandomization as opposed to some earlier derandomizations of the Goemans-Williamson algorithm, which are instance-specific (e.g., [MH99]).

1.3 Proof Outline

In this section, we give a high level outline of the proof of our invariance principle and contrast it with the works of Mossel et al. [MOO05] and Mossel [Mos08]. The proof proceeds in two steps.

Step One: As in Mossel et al. [MOO05] and Mossel [Mos08], we first use the Lindeberg method (see [PR89]) to prove an invariance principle for smooth functions. By this we mean proving that

$$\left| \mathop{\mathbb{E}}_{X \in \{-1,1\}^n} [\Psi(\ell_1(X), \ldots, \ell_k(X))] - \mathop{\mathbb{E}}_{Y \in \mathcal{N}^n} [\Psi(\ell_1(Y), \ldots, \ell_k(Y))] \right| \leq \gamma, \tag{1.1}$$

where $\ell_1, \ldots, \ell_k$ are linear functions (corresponding to the normals of the faces of the k-polytope) and Ψ is a smoothing function. The value γ will depend on k, the coefficients of the ℓ_p's and the derivatives of Ψ. The function Ψ is often called a "test" function and is smooth if there is a uniform bound on its fourth derivative. Notice here that Ψ maps $\mathbb{R}^k$ to $\mathbb{R}$; in [MOO05], they were concerned with the value $\Psi(Q(X))$ for a low-degree polynomial Q and a univariate test function Ψ.

At this point, we could use the k-wise product of a test function constructed by Mossel et al. to approximate the logical AND function. Further, Mossel provides a very general framework for multivariate test functions and gives bounds for the overall error incurred by the hybrid argument. Here we run into our first difficulty: the standard hybrid argument as used by Mossel et al. and Mossel results in a bad dependence on the coefficients of the ℓ_p's. In particular, the resulting error term is not small even for polytopes formed by the intersection of regular halfspaces.

To solve this problem, we use a non-standard hybrid argument that groups the input variables into blocks. We observe that it is irrelevant in which order we replace X_i's with Y_i's – in fact a random order would suffice. Further, we can group the X_i's into blocks and proceed blockwise with the hybrid argument. To implement this intuition, we partition $[n]$ randomly into a set of blocks and replace all the X_i's within a block by the corresponding Y_i's one block at a time. Proceeding in this fashion with a random partitioning has a "smoothing effect" on the coefficients of the linear functions resulting in a much better bound on the error in terms of the coefficients.

Roughly speaking, if ℓ_{pi} denotes the i'th coefficient of ℓ_p, then the standard hybrid arguments of [PR89], [MOO05], [Mos08] incur an error proportional to $\sum_{i \in [n]} \left(\max_{p \in [k]} |\ell_{pi}|^4 \right)$, which can be as large as $\Omega(k)$ even for regular functions ℓ_p. In contrast, our *randomized-blockwise-hybrid* argument only suffers an error of $(\log k) \cdot \max_{p \in [k]} \sum_i |\ell_{pi}|^4$, which is small for regular functions. It turns out that in the above analysis, we can choose the random partitioning into blocks in a $\Theta(\log k)$-wise independent manner, instead of uniformly at random, and this is crucial for our PRG constructions.

Step Two: Given the above invariance principle for smooth functions, we now aim to translate the closeness in expectation for smooth functions to closeness in cdf distance. Here the smoothness of the test function Ψ becomes important, and we run into our second problem: the natural choice of test function Ψ (the multivariate version of the test function from Mossel et al.) leads to an error bound on the order of k, rather than $\mathrm{poly}(\log k)$. To get around this problem, we first observe that in Mossel's proof of the multivariate invariance principle as in our *randomized-blockwise-hybrid* argument, it suffices to bound the 'l_1-norm' of the fourth derivative $\sup_{x \in \mathbb{R}^k} (\sum_{p,q,r,s \in [k]} |\partial_p \partial_q \partial_r \partial_s \Psi(x)|)$, instead of uniformly bounding the fourth derivative $\sup_{x \in \mathbb{R}^k, p,q,r,s \in [k]} (|\partial_p \partial_q \partial_r \partial_s \Psi(x)|)$. Thus, it suffices to obtain a smooth approximation of the AND function for which the former quantity is small. Fortunately for us, we uncovered a beautiful result due to Bentkus [Ben90], who constructs a smooth approximation of the AND function with precisely this property.

The final difficulty for translating closeness in expectation as in Equation 1.1 to closeness in cdf distance is to prove that Ψ differs from the characteristic function only on a set of small Gaussian measure. To this end, we show that it suffices to bound the Gaussian measure of l_∞-neighborhoods around the boundary of k-polytopes. For an l_∞-neighborhood of width λ, a union bound would imply Gaussian measure on the order of $k\lambda$. At this point, however, we can apply a nontrivial result due to Nazarov [Naz03] on the Gaussian surface area of k-polytopes to get the much better

bound of $\sqrt{\log k}\,\lambda$. This result of Nazarov was used before by Klivans et al. [KOS08] in the context of learning intersections of halfspaces with respect to Gaussian distributions.

We give an outline of the proofs of the applications of the invariance principle to noise sensitivity and PRGs in the corresponding sections.

1.4 Related Work

As mentioned earlier, the classical Berry-Esséen theorem [Fel71] from probability, a quantitative version of the Central Limit Theorem, gives an invariance principle for the case of a single halfspace (i.e., $k = 1$). More precisely, for any $w \in \mathbb{R}^n$, such that $\|w\| = 1$ and each coefficient of w is at most ε, the Berry-Esséen theorem states that

$$\left| \Pr_{x \in \{-1,1\}^n} [\langle w, x \rangle \geq t] - \Pr_{x \leftarrow \mathcal{N}^n} [\langle w, x \rangle \geq t] \right| \leq O(\varepsilon).$$

Bentkus [Ben03] proves a multidimensional Berry-Esséen theorem for sums of vector-valued random variables each with identity covariance matrix, whose error term depends on the Gaussian surface area of the test set. Although his paper deals with topics related to our work, his result seems to have no implications in our setting.

There is a long history of research on approximately counting the number of solutions to integer programs, especially with regard to contingency tables [JS97, CD03]. However, not much is known in terms of *deterministic* algorithms, and we believe that our deterministic quasi-polynomial time algorithms for dense covering problems and dense set cover instances is the first result of its kind.

Regarding contingency tables, Dyer [Dye03] gave a deterministic, relative-error approximation for counting solutions to contingency tables that runs in time exponential in the number of rows. In contrast, we obtain an algorithm that runs in quasi-polynomial time in the number of rows (however, we do not give a relative-error approximation). Although not stated explicitly before, it is easy to see that the pseudorandom generator for small space machines of Impagliazzo et al. [INW94] yields a deterministic algorithm for counting $n \times k$ contingency tables with additive error at most ε and run time $2^{O(\log^2(nk/\varepsilon))}$. This is incomparable to our algorithm for contingency tables which has run time $2^{(\log n) \cdot \mathrm{poly}(\log k, 1/\varepsilon)}$. In our case, we obtain a polynomial-time, black-box derandomization for contingency tables with a constant number of rows (for $\varepsilon = O(1)$).

For PRGs for intersections of halfspaces, recently Gopalan et al. [GOWZ09] and Diakonikolas et al. [DKN09] gave results incomparable to ours. Gopalan et al. give generators for arbitrary intersections of k halfspaces with seed length linear in k but logarithmic in $1/\delta$. Diakonikolas et al. show that bounded independence fools intersections of quadratic threshold functions and in particular, get generators with seed length $O((\log n) \cdot \mathrm{poly}(k, 1/\varepsilon))$ fooling intersections of k halfspaces. Due to the at least linear dependence on k, the results of the above works do not yield good algorithms for counting solutions to integer programs, as in this setting k is typically large (e.g., $\mathrm{poly}(n)$).

2 Notation and Preliminaries

We use the following notation.

1. For $W \in \mathbb{R}^{n \times k}$, $\theta \in \mathbb{R}^k$, $\mathcal{K}(W, \theta)$ denotes the polytope $\mathcal{K}(W, \theta) = \{x : W^T x \leq \theta\}$. We say a polytope $\mathcal{K}(W, \theta)$ as above has k faces.

2. Unless stated otherwise, we work with the same polytope $\mathcal{K}(W,\theta)$ and assume that the columns of the matrix W have norm one. We often shorten $\mathcal{K}(W,\theta)$ to $\mathcal{K}$ if W,θ are clear from context. We assume that $k \geq 2$.

3. For $A \in \mathbb{R}^{m_1 \times m_2}$, A^T denotes the transpose of A and for $p \in [m_2]$, A^p denotes the p'th column of A.

4. The all ones vector in $\mathbb{R}^k$ is denoted by $\mathsf{1}_k$.

5. For $u \in \mathbb{R}^k$, define rectangle $\mathsf{Rect}(u) = (-\infty, u_1] \times (-\infty, u_2] \times \cdots \times (-\infty, u_k]$. Note that $x \in \mathcal{K}(W,\theta)$ if and only if $W^T x \in \mathsf{Rect}(\theta)$.

6. $\mathcal{N}^n$ (where $\mathcal{N} = \mathcal{N}(0,1)$) denotes the standard multivariate spherical Gaussian distribution over $\mathbb{R}^n$ with mean 0 and identity covariance matrix.

7. For a smooth function $\psi : \mathbb{R}^k \to \mathbb{R}$, let $\|\psi^{(4)}\|_1 = \sup\{ \sum_{p,q,r,s \in [k]} | \partial_p \partial_q \partial_r \partial_s \psi(a_1, \ldots, a_k) | : (a_1, \ldots, a_k) \in \mathbb{R}^k \}$.

8. We denote all universal constants by c, C, even when we have in mind different constants in the same equation.

Definition 2.1 (regularity)**.** *A vector $u \in \mathbb{R}^n$ is ε-regular if $\sum_i u_i^4 \leq \varepsilon^2 \|u\|^2$. A matrix $W \in \mathbb{R}^{n \times k}$ is ε-regular if every column of W is ε-regular. A polytope $\mathcal{K} = \mathcal{K}(W,\theta)$ is ε-regular if W is ε-regular* [1]*.*

The main results of this paper are applicable to a large class of product distributions that satisfy the following two properties.

Definition 2.2 (proper distributions)**.** *A distribution μ over $\mathbb{R}$ is proper if for $X \leftarrow \mu$, $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = 1$ and $\mathbb{E}[X^3] = 0$.*

Definition 2.3 (hypercontractive distributions)**.** *A distribution μ over $\mathbb{R}$ is hypercontractive, if there exists a constant c such that the following holds. For any m, vector $u \in \mathbb{R}^m$, and any $q \geq 2$,*

$$\left(\mathop{\mathbb{E}}_{X \leftarrow \mu^m} [\, |\langle u, X \rangle|^q \,] \right)^{1/q} \leq c\sqrt{q} \left(\mathop{\mathbb{E}}_{X \leftarrow \mu^m} [\, |\langle u, X \rangle|^2 \,] \right)^{1/2}.$$

Two important examples of distributions that are proper and hypercontractive are the uniform distribution over the hypercube $\{1,-1\}^n$ and the multivariate spherical Gaussian $\mathcal{N}^n$.

We also use the following hypercontractivity inequality for degree d multilinear polynomials over the hypercube (see [O'D08]).

Lemma 2.4 ($(2,q)$-hypercontractivity)**.** *For any $q \geq 2$ and any degree d multilinear polynomial $P : \{1,-1\}^n \to \mathbb{R}$,*

$$\left(\mathop{\mathbb{E}}_{x \in_u \{1,-1\}^n} [\, |P(x)|^q \,] \right)^{1/q} \leq q^{d/2} \left(\mathop{\mathbb{E}}_{x \in_u \{1,-1\}^n} [\, |P(x)|^2 \,] \right)^{1/2}.$$

[1] "Regular polytopes" have a different meaning in combinatorics, but for the purpose of this paper, we will abuse notation and say a polytope is ε-regular if it is formed by the intersection of ε-regular halfspaces as in Definition 2.1.

3 Invariance Principle for Polytopes

Our main invariance principle for polytopes $\mathcal{K}(W,t)$ is as follows:

Theorem 3.1 (invariance principle for polytopes)**.** *For any proper and hypercontractive distribution μ over $\mathbb{R}$ there exists a constant C such that the following holds. For any ε-regular k-polytope $\mathcal{K}$,*

$$\left| \Pr_{X \leftarrow \mu^n}[X \in \mathcal{K}] - \Pr_{Y \leftarrow \mathcal{N}^n}[Y \in \mathcal{K}] \right| \leq C \left(\log^{8/5} k\right) \left(\varepsilon \log(1/\varepsilon)\right)^{1/5}. \tag{3.1}$$

The proof of the theorem can be divided into three parts.

1. We establish an invariance principle for smooth functions on polytopes (Theorem 3.2) using an extension of Lindeberg's method; Section 4 is devoted to proving this part.
2. We prove that for random variables A, B over $\mathbb{R}^k$, closeness with respect to smooth functions and anti-concentration bounds for one of the variables imply closeness with respect to rectangles (Lemma 3.3). To do so, we use a nontrivial result of Bentkus [Ben90] on smooth approximations for the l_∞ norm.
3. We use a result of Nazarov [Naz03] on Gaussian surface area of polytopes to bound the Gaussian measure of "l_∞-neighborhoods" of polytopes in $\mathbb{R}^n$ (Lemma 3.4).

We begin by stating an *invariance principle for smooth functions* $\psi : \mathbb{R}^k \to \mathbb{R}$. The proof is involved, making use of the randomized-blockwise-hybrid argument alluded to in the introduction. For clarity we present the proof in the next section (Section 4).

Theorem 3.2 (invariance principle for smooth functions)**.** *For any proper and hypercontractive distribution μ over $\mathbb{R}$ there exists a constant C such that the following holds. For any ε-regular W and smooth function $\psi : \mathbb{R}^k \to \mathbb{R}$,*

$$\left| \mathop{\mathbb{E}}_{X \leftarrow \mu^n}[\psi(W^T X)] - \mathop{\mathbb{E}}_{Y \leftarrow \mathcal{N}^n}[\psi(W^T Y)] \right| \leq C \|\psi^{(4)}\|_1 \left(\log^3 k\right) \left(\varepsilon \log(1/\varepsilon)\right).$$

The following lemma shows that for two random variables A, B over $\mathbb{R}^k$, closeness with respect to smooth functions and *anti-concentration bounds* for the variable B imply closeness with respect to rectangles. Note that to use the lemma we do not need anti-concentration bounds for the random variable A.

Lemma 3.3 (smooth approximation of AND)**.** *Let A, B be two random variables over $\mathbb{R}^k$ satisfying the following conditions:*

- *For all smooth functions $\psi : \mathbb{R}^k \to \mathbb{R}$, $|\mathbb{E}[\psi(A)] - \mathbb{E}[\psi(B)]| \leq \Delta \|\psi^{(4)}\|_1$.*
- *For a function $g_k : [0,1] \to [0,1]$ the following holds:*
  for all $\lambda \in [0,1]$, $\sup_{\theta \in \mathbb{R}^k} \left(\Pr[\, B \in \mathsf{Rect}(\theta + \lambda 1_k) \setminus \mathsf{Rect}(\theta)\,]\right) \leq g_k(\lambda)$.

Then, for all $\theta \in \mathbb{R}^k$, $0 < \lambda < 1$, $|\Pr[A \in \mathsf{Rect}(\theta)] - \Pr[B \in \mathsf{Rect}(\theta)]| \leq C\Delta \log^3 k/\lambda^4 + C g_k(\lambda)$.

Finally, we use the following anti-concentration bound that follows from Nazarov's estimate on the Gaussian surface area of polytopes [Naz03]:

Lemma 3.4 (anti-concentration bound for l_∞-neighborhood of rectangles)**.** *For $0 < \lambda < 1$,*

$$\Pr_{x \leftarrow \mathcal{N}^n} \left[W^T x \in \mathsf{Rect}(\theta) \setminus \mathsf{Rect}(\theta - \lambda 1_k) \right] = O(\lambda \sqrt{\log k}).$$

We first prove Theorem 3.1 using the above three results and then prove Lemmas 3.3 and 3.4 in Sections 3.1 and 3.2. Theorem 3.2 is then proved in Section 4.

Proof of Theorem 3.1. Let $X \leftarrow \mu^n$, $Y \leftarrow \mathcal{N}^n$ and let random variables $A = W^T X$, $B = W^T Y$. Then, by Lemma 3.4 and Theorem 3.2,

$$\Pr\left[B \in \mathbb{R}(\theta + \lambda 1_k) \setminus \mathbb{R}(\theta)\right] \le C\sqrt{\log k}\,\lambda \text{ and } \left|\mathbb{E}[\psi(A)] - \mathbb{E}[\psi(B)]\right| \le C\,(\log^3 k)\,\varepsilon\,\log(1/\varepsilon)\,\|\psi^{(4)}\|_1,$$

where $\psi : \mathbb{R}^k \to \mathbb{R}$ is any smooth function, $\theta \in \mathbb{R}^k$ and $\lambda \in (0,1)$. Therefore, by Lemma 3.3, for $\theta \in \mathbb{R}^k$,

$$\left|\Pr\left[A \in \mathsf{Rect}(\theta)\right] - \Pr\left[B \in \mathsf{Rect}(\theta)\right]\right| \le C\,(\log^6 k)\,\log(1/\varepsilon)\varepsilon/\lambda^4 + C\sqrt{\log k}\,\lambda.$$

The theorem now follows by setting $\lambda = (\log^{11/10} k)\,(\varepsilon \log(1/\varepsilon))^{1/5}$. ∎

3.1 Smooth approximation of AND

We now prove Lemma 3.3. For this, we use the following nontrivial result of Bentkus [Ben90] on smooth approximations for the l_∞ norm.

Theorem 3.5 (Bentkus [Ben90])**.** *For every $\alpha > 0$ and $0 < \lambda < 1$, there exists a function $\psi \equiv \psi_{\alpha,\lambda} : \mathbb{R}^k \to \mathbb{R}$ such that $\|\psi^{(4)}\|_1 \le C \log^3 k/\lambda^4$ and*

$$\psi(a) = \begin{cases} 1 & \text{if } \|a\|_\infty \le \alpha \\ 0 & \text{if } \|a\|_\infty > \alpha + \lambda \\ \in [0,1] & \text{otherwise} \end{cases}.$$

Corollary 3.6. *For all $u \in \mathbb{R}^k$, $0 < \lambda < 1$, $T > \|u\|_\infty$, there exists a function $\psi \equiv \psi_{u,\lambda,T} : \mathbb{R}^k \to \mathbb{R}$ such that $\|\psi^{(4)}\|_1 \le C \log^3 k/\lambda^4$ and*

$$\psi(a) = \begin{cases} 1 & \text{if } \forall l \in [k], -T + u_l \le a_l \le u_l \\ 0 & \text{if } \exists l \in [k], a_l > u_l + \lambda \\ \in [0,1] & \text{otherwise} \end{cases}.$$

Proof. Let $\psi_{T/2,\lambda}$ be the function from Theorem 3.5 with $\alpha = T/2$. Define $\psi \equiv \psi_{u,\lambda,T} : \mathbb{R}^k \to \mathbb{R}$ by

$$\psi_{u,\lambda,T}(a_1, \ldots, a_k) = \psi_{T/2,\lambda}(a_1 + T/2 - u_1, a_2 + T/2 - u_2, \ldots, a_k + T/2 - u_k).$$

It is easy to check that ψ satisfies the conditions of the theorem. ∎

Proof of Lemma 3.3. Fix $\theta \in \mathbb{R}^k$, $0 < \lambda < 1$. Choose $T \in \mathbb{R}$ large enough so that $T > \|\theta\|_\infty$, $\Pr[\|A\|_\infty \ge T] < \Delta$ and $\Pr[\|B\|_\infty \ge T] < \Delta$. Let $\psi : \mathbb{R}^k \to \mathbb{R}$ be the function obtained from applying Corollary 3.6 to θ, λ, T. Then,

$$\left|\Pr\left[A \in \mathsf{Rect}(\theta)\right] - \Pr\left[A \in \mathsf{Rect}_T(\theta)\right]\right| \le \Delta, \quad \left|\Pr\left[B \in \mathsf{Rect}(\theta)\right] - \Pr\left[B \in \mathsf{Rect}_T(\theta)\right]\right| \le \Delta, \qquad (3.2)$$

where $\mathsf{Rect}_T(\theta) = [-T+\theta_1, \theta_1] \times [-T+\theta_2, \theta_2] \times \cdots \times [-T+\theta_k, \theta_k] \subseteq \mathbb{R}^k$. Observe that from the definition of ψ in Corollary 3.6 and Equation 3.2

$$\mathsf{Pr}\left[A \in \mathsf{Rect}(\theta)\right] \leq \mathbb{E}[\psi(A)] + \Delta \leq \mathbb{E}[\psi(B)] + \Delta\|\psi^{(4)}\|_1 + \Delta.$$

Similarly,

$$\mathbb{E}[\psi(B)] \leq \mathsf{Pr}[B \in \mathsf{Rect}(\theta + \lambda 1_k)] = \mathsf{Pr}[B \in \mathsf{Rect}(\theta)] + \mathsf{Pr}[B \in \mathsf{Rect}(\theta + \lambda 1_k) \setminus \mathsf{Rect}(\theta)] \leq \mathsf{Pr}[B \in \mathsf{Rect}(\theta)] + g_k(\lambda),$$

where the last inequality follows from the definition of g_k. Combining the above two equations we get

$$\mathsf{Pr}[A \in \mathsf{Rect}(\theta)] \leq \mathsf{Pr}[B \in \mathsf{Rect}(\theta)] + 2\Delta\|\psi^{(4)}\|_1 + g_k(\lambda) \leq \mathsf{Pr}[B \in \mathsf{Rect}(\theta)] + \frac{C\Delta \log^3 k}{\lambda^4} + g_k(\lambda).$$

Proceeding similarly for the function $\psi_L : \mathbb{R}^k \to \mathbb{R}$ obtained by applying Corollary 3.6 to $t - \lambda 1_k, \lambda, T$, we get

$$\mathsf{Pr}\left[A \in \mathsf{Rect}(\theta)\right] \geq \mathsf{Pr}\left[B \in \mathsf{Rect}(\theta)\right] - \frac{C\Delta \log^3 k}{\lambda^4} - g_k(\lambda).$$

Therefore,

$$\left|\mathsf{Pr}\left[A \in \mathsf{Rect}(\theta)\right] - \mathsf{Pr}\left[B \in \mathsf{Rect}(\theta)\right]\right| \leq \frac{C\Delta \log^3 k}{\lambda^4} + g_k(\lambda).$$

□

3.2 Anti-concentration bound for l_∞-neighborhood of rectangles

Lemma 3.4 follows straightforwardly from the following result of Nazarov [Naz03]. For a convex body $K \subseteq \mathbb{R}^n$ with boundary ∂K, let $\Gamma(K)$ denote the Gaussian surface area of K defined by

$$\Gamma(K) = \int_{y \in \partial K} e^{\frac{-\|y\|^2}{2}} d\sigma(y),$$

where $d\sigma(y)$ denotes the surface element at $y \in \partial K$.

Theorem 3.7 (Nazarov (see [KOS08, Theorem 20]))**.** *For a polytope $\mathcal{K}$ with at most k faces, $\Gamma(\mathcal{K}) \leq C\sqrt{\log k}$.*

Proof of Lemma 3.4. Consider an increasing (under set inclusion) family of polytopes $\mathcal{K}_\rho$ for $0 \leq \rho \leq \lambda$ such that $\mathcal{K}_0 = \{x : W^T x \in \mathsf{Rect}(\theta - \lambda 1_k)\}$ and $\mathcal{K}_\lambda = \{x : W^T x \in \mathsf{Rect}(\theta)\}$. Then,

$$\mathop{\mathsf{Pr}}_{x \leftarrow \mathcal{N}^n}\left[W^T x \in \mathsf{Rect}(\theta) \setminus \mathsf{Rect}(\theta - \lambda 1_k)\right] = \int_{\rho=0}^{\lambda} \Gamma(\mathcal{K}_\rho) d\rho \leq C\sqrt{\log k}\,\lambda,$$

where the last inequality follows from Theorem 3.7. □

4 Invariance principle for Smooth Functions over Polytopes

We now prove Theorem 3.2. The proof of the theorem is based on Lindeberg's method. Let $t = 1/\varepsilon$ and let $\mathcal{H} = \{h : [n] \to [t]\}$ be a family of $(2\log k)$-wise independent functions. That is, for all $I \subseteq [n], |I| \leq 2\log k$ and $b \in [t]^I$, $\mathsf{Pr}_{h\in_u\mathcal{H}}[\forall i \in I,\ h(i) = b_i\,] = \frac{1}{t^{|I|}}$.

We remark that to prove Theorem 3.2 we could take the hash family to be the set of all functions. However, we work with a $(2\log k)$-wise independent family as the analysis is no more complicated and we need to work with such hash families while constructing pseudorandom generators. For $S \subseteq [n]$, let W_S be the matrix formed by the rows of W with indices in S. Define

$$\mathcal{H}(W) \stackrel{\text{def}}{=} \sum_{i=1}^{t} \left(\mathop{\mathbb{E}}_{h} \left[\sum_{p=1}^{k} \|W^p_{h^{-1}(i)}\|^{4\log k} \right] \right)^{1/\log k}.$$

Theorem 3.2 follows immediately from the following two lemmas.

Lemma 4.1. *For ε-regular W, $\mathcal{H}(W) \leq C\,\log k\,(\varepsilon\,\log(1/\varepsilon))$.*

Lemma 4.2. *For any smooth function $\psi : \mathbb{R}^k \to \mathbb{R}$,*

$$\left| \mathop{\mathbb{E}}_{X\leftarrow\mu^n} \left[\psi(W^T X)\right] - \mathop{\mathbb{E}}_{Y\leftarrow\mathcal{N}^n} \left[\psi(W^T Y)\right] \right| \leq 4\,(\log^2 k)\,\mathcal{H}(W)\|\psi^{(4)}\|_1.$$

Proof of Lemma 4.1. Fix a $l \in [t]$, $p \in [k]$. For $i \in [n]$, let X_i be the indicator random variable that is 1 if $h(i) = l$ and 0 otherwise. Then, $\mathsf{Pr}[X_i = 1] = 1/t$ and the variables $X_1, \ldots, X_n$ are $(2\log k)$-wise independent. Further,

$$Z'_p \equiv \|W^p_{|h^{-1}(l)}\|^2 = \sum_{i=1}^{n} W^2_{ip} X_i.$$

Let Y_i be i.i.d indicator random variables with $\mathsf{Pr}[Y_i = 1] = 1/t$ and let $Z_p = \sum_{i=1}^n W^2_{ip} Y_i$. Observe that Z'_p and Z_p have identical d'th moments for $d \leq 2\log k$. Moreover, by Hoeffding's inequality applied to Z_p, for any $\gamma > 0$,

$$\mathsf{Pr}\left[\left| Z_p - \frac{1}{t} \right| \geq \gamma \right] \leq 2\exp\left(-\frac{2\gamma^2}{\sum_{i=1}^n W^4_{ip}} \right) \leq 2\exp\left(-\frac{2\gamma^2}{\varepsilon^2} \right) = 2\exp(-2t^2\gamma^2).$$

The above tail bound for Z_p implies strong bounds on the moments of Z_p by standard arguments. Setting $\gamma = \sqrt{2\log k\,\log t}/t$ in the above equation, we get

$$\mathsf{Pr}\left[|Z_p| \geq \frac{\sqrt{3\log k\,\log t}}{t} \right] \leq \frac{1}{t^{2\log k}}.$$

Therefore, from the above equation and the fact that $Z_p \leq 1$

$$\begin{aligned} \mathbb{E}[Z_p^{2\log k}] &\leq \frac{(3\log k\,\log t)^{\log k}}{t^{2\log k}} + \mathsf{Pr}\left[|Z_p| \geq \frac{\sqrt{3\log k\,\log t}}{t} \right] \\ &\leq \frac{(4\log k\,\log t)^{\log k}}{t^{2\log k}}. \end{aligned}$$

Therefore,

$$\mathop{\mathbb{E}}_{h\in_u\mathcal{H}}\left[\,\|W^p_{|h^{-1}(l)}\|^{4\log k}\,\right]=\mathbb{E}\left[\,(Z'_p)^{2\log k}\right]=\mathbb{E}\left[\,Z_p^{2\log k}\,\right]\le\frac{(4\log k\log t)^{\log k}}{t^{2\log k}}.$$

Therefore, from the definition of $\mathcal{H}(W)$ and the above equation,

$$\mathcal{H}(W)=\sum_{i=1}^{t}\left(\sum_{p=1}^{k}\mathop{\mathbb{E}}_{h}\left[\,\|W^p_{h^{-1}(i)}\|^{4\log k}\right]\right)^{1/\log k}\le t\frac{4\log k\log t}{t^2}=4(\log k)(\varepsilon\log(1/\varepsilon)).$$

□

The proof of Lemma 4.2 uses a blockwise hybrid argument and careful applications of hypercontractivity as sketched in the proof outline in the introduction.

We use the following form of the standard Taylor series expansion. For a smooth function $\psi:\mathbb{R}^k\to\mathbb{R}$, $x\in\mathbb{R}^k$ and $p_1,\ldots,p_r\in[k]$, let $\partial_{p_1,\ldots,p_r}\psi(x)=\partial_{p_1}\partial_{p_2}\cdots\partial_{p_r}\psi(x)$. For indices $p_1,\ldots,p_r\in[k]$, let $(p_1,\ldots,p_r)!=s_1!s_2!\ldots s_k!$, where, for $l\in[k]$, s_l denotes the number of occurrences of l in $(p_1,\ldots,p_r)$.

Fact 4.3 (Multivariate Taylor's Theorem)**.** *For any smooth function $\psi:\mathbb{R}^k\to\mathbb{R}$, and $x,y\in\mathbb{R}^k$,*

$$\psi(x+y)=\psi(x)+\sum_{p\in[k]}\partial_p\psi(x)\,y_p+\sum_{p,q\in[k]}\frac{1}{(p,q)!}\,\partial_{p,q}\psi(x)\,y_py_q+\sum_{p,q,r\in[k]}\frac{1}{(p,q,r)!}\,\partial_{p,q,r}\psi(x)\,y_py_qy_r+\mathsf{err}(x,y),$$

where $|\mathsf{err}(x,y)|\le\|\psi^{(4)}\|_1\cdot\max_{p\in[k]}|y_p|^4$.

Proof of Lemma 4.2. Let $\overline{X}\leftarrow\mu^n$ and $\overline{Y}\leftarrow\mathcal{N}^n$. We first partition $[n]$ into blocks using a random hash function $h\in_u\mathcal{H}$ and then use a blockwise-hybrid argument. Fix a hash function $h\in\mathcal{H}$. View $\overline{X}$ as $X^1,\ldots,X^t$, where each $X^l=\overline{X}_{h^{-1}(l)}$ is chosen independently and uniformly from $\mu^{|h^{-1}(l)|}$. Similarly, view $\overline{Y}$ as $Y^1,\ldots,Y^t$ where each $Y^l=\overline{Y}_{h^{-1}(l)}$ is chosen independently and uniformly from $\mathcal{N}^{|h^{-1}(l)|}$. We prove the claim via a hybrid argument where we replace the blocks $X^1,\ldots,X^t$ with $Y^1,\ldots,Y^t$ one at a time.

For $0\le i\le t$, let Z^i be the distribution with $Z^i_{|h^{-1}(j)}=X^j$ for $i<j\le t$ and $Z^i_{|h^{-1}(j)}=Y^j$ for $1\le j\le i$. Then, Z^0 is distributed as μ^n and Z^t is distributed as $\mathcal{N}^n$. For $l\in[t]$, let

$$h(W,l)=\left(\sum_{p=1}^{k}\|W^p_{h^{-1}(l)}\|^{4\log k}\right)^{1/\log k}.$$

Claim 4.4. *For $1\le l\le t$, and fixed $h\in\mathcal{H}$,*

$$\left|\mathop{\mathbb{E}}_{X,Y}\left[\psi(W^TZ^l)\right]-\mathop{\mathbb{E}}_{X,Y}\left[\psi(W^TZ^{l-1})\right]\right|\le C\,\log^2k\,\|\psi^{(4)}\|_1\,h(W,l).$$

Proof. Without loss of generality, suppose that $h^{-1}(l)=\{1,\ldots,m\}$. Note that Z^l,Z^{l-1} have the same random variables in positions $m+1,\ldots,n$. Let $Z^{l-1}=(X_1,\ldots,X_m,Z_{m+1},\ldots,Z_n)$ and $Z^l=(Y_1,\ldots,Y_m,Z_{m+1},\ldots,Z_n)$ where $(X_1,\ldots,X_m)$ is uniform over μ^m and $(Y_1,\ldots,Y_m)$ is uniform over $\mathcal{N}^m$. Note that $(Z_{m+1},\ldots,Z_n)$ is independent of $(X_1,\ldots,X_m)$, $(Y_1,\ldots,Y_m)$.

Let $W_1\in\mathbb{R}^{m\times k}$ be the matrix formed by the first m rows of W and similarly let $W_2\in\mathbb{R}^{(n-m)\times k}$ be the matrix formed by the last $n-m$ rows of W. Lastly, let $V=W_2^T(Z_{m+1},\ldots,Z_n)$ and U be

one of $X = (X_1, \ldots, X_m)$ or $Y = (Y_1, \ldots, Y_m)$. Now, by using a Taylor expansion of ψ at V as in Fact 4.3,

$$\begin{aligned}\psi(W^T(U_1, \ldots, U_m, Z_{m+1}, \ldots, Z_n)) &= \psi(\, W_1^T U + V\,) \\ &= \psi(V) + \sum_{p\in[k]} \partial_p \psi(V)\, \langle W_1^p, U\rangle + \sum_{p,q\in[k]} \frac{1}{(p,q)!}\, \partial_{p,q}\psi(V)\, \langle W_1^p, U\rangle\, \langle W_1^q, U\rangle \\ &\quad + \sum_{p,q,r\in[k]} \frac{1}{(p,q,r)!}\, \partial_{p,q,r}\psi(V)\, \langle W_1^p, U\rangle\, \langle W_1^q, U\rangle\, \langle W_1^r, U\rangle + \mathsf{err}(V, W_1^T U). \quad (4.1)\end{aligned}$$

Now, using the fact that $\|z\|_\infty \le \|z\|_{\log k}$ for $z \in \mathbb{R}^k$,

$$\left|\mathsf{err}(V, W_1^T U)\right| \le \|\psi^{(4)}\|_1 \cdot \max_{p\in[k]} |\langle W_1^p, U\rangle|^4 \le \|\psi^{(4)}\|_1 \left(\sum_{p=1}^k |\langle W_1^p, U\rangle|^{4\log k}\right)^{1/\log k}. \quad (4.2)$$

Now, by hypercontractivity of μ,

$$\begin{aligned}\mathop{\mathbb{E}}_X \left[\left(\sum_{p=1}^k |\langle W_1^p, X\rangle|^{4\log k}\right)^{1/\log k}\right] &\le \left(\mathop{\mathbb{E}}_X \left[\sum_{p=1}^k |\langle W_1^p, X\rangle|^{4\log k}\right]\right)^{1/\log k} && \text{(by power-mean inequality)} \\ &= \left(\sum_{p=1}^k \mathop{\mathbb{E}}_X \left[|\langle W_1^p, X\rangle|^{4\log k}\right]\right)^{1/\log k} \\ &\le \left(\sum_{p=1}^k (C\log k)^{2\log k}\, \|W_1^p\|^{4\log k}\right)^{1/\log k} && \text{(by hypercontractivity of } \mu) \\ &\le C(\log^2 k)\, h(W, l). && (4.3)\end{aligned}$$

Similarly, by hypercontractivity of $\mathcal{N}$,

$$\mathop{\mathbb{E}}_Y \left[\left(\sum_{p=1}^k |\langle W_1^p, Y\rangle|^{4\log k}\right)^{1/\log k}\right] \le C(\log^2 k)\, h(W, l). \quad (4.4)$$

Since μ is proper, for any $u^1, u^2, u^3 \in \mathbb{R}^m$,

$$\mathbb{E}[\langle u^1, X\rangle] = \mathbb{E}[\langle u^1, Y\rangle], \quad \mathbb{E}[\langle u^1, X\rangle\, \langle u^2, X\rangle] = \mathbb{E}[\langle u^1, Y\rangle\, \langle u^2, Y\rangle]$$

$$\mathbb{E}[\langle u^1, X\rangle\, \langle u^2, X\rangle\, \langle u^3, X\rangle] = \mathbb{E}[\langle u^1, Y\rangle\, \langle u^2, Y\rangle\, \langle u^3, Y\rangle].$$

From the above equations, Equations (4.1), (4.2), (4.3), (4.4) and the fact that X, Y, V are independent of one another, it follows that $\left|\mathbb{E}\left[\psi(W^T Z^l) - \psi(W^T Z^{l-1})\right]\right| \le C(\log^2 k)\, \|\psi^{(4)}\|_1 h(W, l)$. □

Lemma 4.2 now follows from the above claim, summing from $l = 1, \ldots, t$, and taking expectation with respect to $h \in_u \mathcal{H}$. □

5 Noise Sensitivity of Intersections of Regular Halfspaces

We now describe how our invariance principle yields a bound on the average and noise sensitivity of intersections of regular halfspaces. We begin by defining the (Boolean) noise sensitivity of a Boolean function:

Definition 5.1 (noise sensitivity)**.** *Let f be a Boolean function $f : \{1,-1\}^n \to \{1,-1\}$. For any $\delta \in (0,1)$, let X be a random element of the hypercube $\{1,-1\}^n$ and Z a δ-perturbation of X defined as follows: for each i independently, Z_i is set to X_i with probability $1-\delta$ and $-X_i$ with probability δ. The noise sensitivity of f, denoted $\mathbb{NS}_\delta(f)$, for noise δ is then defined as follows:* $\mathbb{NS}_\delta(f) = \Pr[f(X) \neq f(Z)]$.

Let $f^1, \ldots, f^k : \{1,-1\}^n \to \{1,-1\}$ be halfspaces with $f^p(x) = \mathsf{sign}(\langle W^p, x\rangle - \theta_p)$ and let $f^{\wedge k} : \{1,-1\}^n \to \{1,-1\}$ be their intersection, $f^{\wedge k} = f^1 \wedge f^2 \wedge \ldots \wedge f^k$.

Theorem 5.2. *For $f^{\wedge k}$ ε-regular, $\mathbb{NS}_\delta(f^{\wedge k}) \leq C(\log^{1.6}(k/\delta))\,(\varepsilon^{1/6} + \delta^{1/2})$.*

We prove the theorem by first reducing bounding noise sensitivity of $f^{\wedge k}$ to bounding the Boolean volume of l_∞-neighborhoods of polytopes. We then use our invariance principle, Theorem 3.1, to prove the required bounds on the Boolean volume of boundaries of polytopes.

As mentioned before, the above theorem implies a $n^{\log^{O(1)} k}$ algorithm for learning intersections of regular halfspaces in the agnostic model for any constant error rate.

We use the following tail bound that follows from Pinelis's subgaussian tail estimates [Pin94].

Fact 5.3. *There exist absolute constants $c_1, c_2 > 0$ such that all $w \in \mathbb{R}^m$, $t > 0$,*

$$\Pr_{x \in_u \{1,-1\}^m}[\,|\langle w, x\rangle| > t\|w\|\,] \leq c_1 \exp(-c_2 t^2).$$

The following claim says that for W ε-regular, random $x \in_u \{1,-1\}^n$, and a δ-perturbation y of x, $W^T x$ is close to $W^T y$ in l_∞ distance.

Claim 5.4. *For $x \in \{1,-1\}^n$, let $y(x)$ be a random δ-perturbation of $y(x)$ of x. Then,*

$$\Pr_{x \in_u \{1,-1\}^n, y(x)}\left[\,\|W^T x - W^T y(x)\|_\infty \geq \lambda\,\right] \leq 2\delta,$$

where $\lambda = C\log(k/\delta)^{1/2}\delta^{1/2} + C\log(k/\delta)^{3/4}\varepsilon^{1/2}$.

Proof. Let $Y = (Y_1, \ldots, Y_n)$ be i.i.d indicator variables with $\Pr[Y_i = 1] = \delta$. Let $S(Y) = \mathit{support}(Y)$. Now, for $p \in [k]$, $\|W^p_{S(Y)}\|^2 = \sum_{i=1}^n W^2_{ip} Y_i$ and $\mathbb{E}[\|W^p_{S(Y)}\|^2] = \delta$. Further, since W is ε-regular, by Hoeffding's inequality, for all $t > 0$,

$$\Pr\left[\,|\|W^p_{S(Y)}\|^2 - \delta| \geq \gamma\right] \leq 2\exp\left(\frac{-2\gamma^2}{\sum_i W^4_{ip}}\right) \leq 2\exp\left(\frac{-2\gamma^2}{\varepsilon^2}\right).$$

Thus, by a union bound

$$\Pr_Y\left[\,\exists p \in [k],\ \|W^p_{S(Y)}\|^2 \geq \delta + 2\sqrt{\log(k/\delta)}\,\varepsilon\,\right] \leq \delta. \tag{5.1}$$

Note that for a fixed Y and sufficiently large C, by Fact 5.3 and a union bound,

$$\Pr_{x \in_u \{1,-1\}^n}\left[\,\exists p \in [k],\ |\langle W^p_{S(Y)}, x_{S(Y)}\rangle| \geq C\sqrt{\log(k/\delta)}\,\|W^p_{S(Y)}\|\,\right] \leq \delta.$$

From Equation 5.1 and the above equation, we get that for a sufficiently large constant C

$$\Pr_{x\in_u\{1,-1\}^n,Y}\left[\exists p\in[k],\ |\langle W^p_{S(Y)},x_{S(Y)}\rangle|\geq C\log(k/\delta)^{1/2}\delta^{1/2}+C\log(k/\delta)^{3/4}\varepsilon^{1/2}\right]\leq 2\delta. \tag{5.2}$$

Now, observe that that for $x\in\{1,-1\}^n$, to generate a δ-perturbation of x, $y(x)$, we can first generate a random Y as above and flip the bits of x in the support of Y. Thus, from Equation 5.2,

$$\begin{aligned}\Pr_{x\in_u\{1,-1\}^n,Y}[\,\exists p\in[k],\ |\langle W^p,x\rangle-\langle W^p,y(x)\rangle|\geq\lambda\,]&=\Pr_{x\in_u\{1,-1\}^n,Y}\left[\exists p\in[k],\ 2\,|\langle W^p_{S(Y)},x_{S(Y)}\rangle|\geq\lambda\right]\\&\leq 2\delta,\end{aligned}$$

where $\lambda=C\log(k/\delta)^{1/2}\delta^{1/2}+C\log(k/\delta)^{3/4}\varepsilon^{1/2}$. Therefore,

$$\Pr_{x\in_u\{1,-1\}^n,Y}[\,\|W^Tx-W^Ty(x)\|_\infty\geq\lambda\,]\leq 2\delta.$$

□

The following claim can be seen as an anti-concentration bound for regular polytopes over the hypercube and could be of use elsewhere.

Claim 5.5. *For ε-regular $W\in\mathbb{R}^{n\times k}$, $\theta\in\mathbb{R}^k$, and $0<\lambda<1$,*

$$\Pr_{x\in_u\{1,-1\}^n}[\,W^Tx\in\mathsf{Rect}(\theta+\lambda\,1_k)\setminus\mathsf{Rect}(\theta-\lambda\,1_k)\,]\leq C(\log^{1.6}k)\,(\varepsilon\,\log(1/\varepsilon))^{1/5}+\sqrt{\log k}\,\lambda.$$

Proof. Follows directly from Theorem 3.1 and Lemma 3.4. □

We can now prove Theorem 5.2.

Proof of Theorem 5.2. Note that for $x,y\in\mathbb{R}^n$, $f^{\wedge k}(x)\neq f^{\wedge k}(y)$ implies that $W^Tx\in\mathsf{Rect}(\theta+\gamma 1_k)\setminus\mathsf{Rect}(\theta-\gamma 1_k)$, where $\gamma=\|W^Tx-W^Ty\|_\infty$. Hence,

$$\begin{aligned}\mathbb{NS}_\delta(f^{\wedge k})&=\Pr_{x\in_u\{1,-1\}^n,Y}[\,f^{\wedge k}(x)\neq f^{\wedge k}(y(x))\,]\\&\leq\Pr_{x\in_u\{1,-1\}^n,Y}[\,f^{\wedge k}(x)\neq f^{\wedge k}(y(x))\,|\,\|W^Tx-W^Ty(x)\|_\infty\leq\lambda\,]+2\delta&&\text{(Claim 5.4)}\\&\leq\Pr_{x\in_u\{1,-1\}^n}[\,W^Tx\in\mathsf{Rect}(\theta+\lambda\,1_k)\setminus\mathsf{Rect}(\theta-\lambda\,1_k)\,]+2\delta\\&\leq C(\log^{1.6}k)\,(\varepsilon\,\log(1/\varepsilon))^{1/5}+\sqrt{\log k}\,\lambda+2\delta.&&\text{(Claim 5.5)}\end{aligned}$$

The theorem now follows. □

6 Pseudorandom Generators for Polytopes

We now prove our main theorems for constructing pseudorandom generators for polytopes with respect to a variety of distributions (Theorems 1.5, 1.6, and 1.7).

The results in this section are based on a recent PRG construction due to Meka and Zuckerman [MZ09] for polynomial threshold functions using the invariance principle of Mossel et al. [MOO05]. A closer look at their construction reveals a general program for constructing PRGs from invariance principles. Given this observation, it is natural to ask if our invariance principle can be used to construct PRGs for regular polytopes. Indeed it can, and we use the Meka and Zuckerman generator but with a different setting of its parameters. The analysis, however, is a little more complicated in our setting (even given our invariance principle) and requires a careful application of hypercontractivity.

6.1 Main Generator Construction

We begin by describing the construction of the PRG we use; it is a slightly modified version of the PRG used by Meka and Zuckerman [MZ09] to fool regular *halfspaces* (i.e., the case $k = 1$).

Give $\delta \in (0,1)$, let $\varepsilon = \Omega(\delta^6/\log^{9.6} k)$ be such that $\log^{1.6} k(\varepsilon \log(1/\varepsilon))^{1/5} = \delta$. Let $t = 1/\varepsilon$ and let $\mathcal{H} = \{h : h : [n] \to [t]\}$ be a $(2\log k)$-wise independent family of hash functions. That is, for all $I \subseteq [n], |I| \leq 2\log k$ and $b \in [t]^I$,

$$\Pr_{h \in_u \mathcal{H}}[\,\forall i \in I,\ h(i) = b_i\,] = \frac{1}{t^{|I|}}.$$

Efficient constructions of hash families $\mathcal{H}$ as above with $|\mathcal{H}| = O(n^{2\log k})$ are known. To avoid some technical issues that can be overcome easily, we assume that every hash function $h \in \mathcal{H}$ is equi-distributed in the following sense: for all $j \in [t]$, $|\{i : h(i) = j\}| = n/t$.

Let $m = n/t$ and let $G_0 : \{0,1\}^s \to \{1,-1\}^m$ generate a $(4\log k)$-wise independent distribution over $\{1,-1\}^m$. That is, for all $I \subseteq [n], |I| \leq 2\log k$ and $b \in \{1,-1\}^I$,

$$\Pr_{x = G_0(z), z \in_u \{0,1\}^s}[\,\forall i \in I,\ x_i = b_i\,] = \frac{1}{2^{|I|}}.$$

Efficient constructions of generators G_0 as above with $s = O(\log k \log n)$ are known [NN93].

Given a hash family and generator G_0 as above, we consider the following generator. Define $G : \mathcal{H} \times (\{0,1\}^s)^t \to \{1,-1\}^n$ by $G(h, z^1, \ldots, z^t) = x$, where $x_{|h^{-1}(i)} = G_0(z^i)$ for $i \in [t]$.

6.2 Pseudorandom Generators for Regular Polytopes

We now argue that the generator G defined in the last section fools regular polytopes and prove Theorem 1.5.

Proof of Theorem 1.5. The bound on the seed length of the generator G follows from the construction. The following statement follows from an argument similar to that of the proof of Theorem 3.2: for any smooth function $\psi : \mathbb{R}^k \to \mathbb{R}$ and ε-regular W,

$$\left| \mathop{\mathbb{E}}_{y \in_u \{0,1\}^r}[\psi(W^T G(y))] - \mathop{\mathbb{E}}_{Y \leftarrow \mathcal{N}^n}[\psi(W^T Y)] \right| \leq C \log^3 k\,(\varepsilon \log(1/\varepsilon))\,\|\psi^{(4)}\|_1. \tag{6.1}$$

Indeed, to observe that Lemma 4.1 holds for any $(2\log k)$-wise independent family of hash functions and the proof of Lemma 4.2 relies only on two key properties of $X \leftarrow \mu^n$: (1) For a fixed hash function h, the blocks $X_{h^{-1}(1)}, X_{h^{-1}(2)}, \ldots, X_{h^{-1}(t)}$ are independent of one another. (2) For a fixed hash function h, and $j \in [t]$, the distribution of each block $X_{h^{-1}(j)}$ satisfies $(2, 2\log k)$-hypercontractivity for all $j \in [t]$. In other words, we used the property that for all $j \in [t]$, $u \in \mathbb{R}^{|h^{-1}(j)|}$,

$$\mathbb{E}[|\langle u, X_{h^{-1}(j)}\rangle|^{4\log k}] \leq (C\log k)^{2\log k}\,\|u\|^{4\log k}. \tag{6.2}$$

Note that X generated according to the generator G satisfies both the above conditions: 1) For a fixed function h, the blocks are independent by definition and 2) the hypercontractivity inequality 6.2 only involves the first $(4\log k)$-moments of the distribution of $X_{h^{-1}(j)}$. As a consequence, inequality 6.2 holds for any $(4\log k)$-wise independent distribution over $\{1,-1\}^{|h^{-1}(j)|}$.

We can now move from closeness in expectation to closeness in cdf distance by an argument similar to the proof of Theorem 3.1, where we use Equation 6.1 instead of Theorem 3.2, to get

$$|\Pr_{y \in_u \{0,1\}^r}[G(y) \in \mathcal{K}] - \Pr_{Y \leftarrow \mathcal{N}^n}[Y \in \mathcal{K}]| \leq \delta.$$

The theorem now follows from the above equation and Theorem 3.1. ∎

6.2.1 Approximate Counting for Integer Programs

The PRG from Theorem 1.5 coupled with enumeration over all possible seeds immediately implies a quasi-polynomial time, deterministic algorithm for approximately counting, within a small additive error, the number of solutions to "regular" $\{0,1\}$-integer programs. It turns out that "regular" integer programs correspond to a broad class of well-studied combinatorial problems. For example, we obtain deterministic, approximate counting algorithms for *dense* set cover problems and $\{0,1\}$-contingency tables. We obtain quasi-polynomial time algorithms even when there are a polynomial number of constraints (or polynomial number of rows in the contingency table setting). As far as we know, there is no prior work giving nontrivial *deterministic* algorithms for counting solutions to integer programs with many constraints.

Here we discuss the case of *dense* set cover instances and remark that we get similar results for the special case of counting contingency tables. Covering integer programs are a fundamental class of integer programs and can be formulated as follows.

$$\begin{aligned} &\min \sum_i X_i \\ \text{s.t. } &\sum_i a_{ij} X_i \geq c_j,\ j = 1, \ldots, k, \\ &X \in \{0,1\}^n, \end{aligned} \tag{6.3}$$

where the coefficients of the constraints a_{ij} and c_j are all non-negative. An important special class of covering integer programs is set cover, which in turn is a generalization of many important problems in combinatorial optimization such as edge cover and multidimensional $\{0,1\}$-knapsack.

In the standard set cover problem, the input is a family of sets $S_1, \ldots, S_n$ over a universe U of size k and an integer t. The goal is to find a subfamily of sets $\mathcal{C}$ such that $|\mathcal{C}| \leq t$ and the union of all the sets in $\mathcal{C}$ equals U. This corresponds to a covering program as above with k constraints and n unknowns from $\{0,1\}$. Call an instance of set cover ε-dense if each element in U appears in at least $1/\varepsilon^2$ of the different sets S_i. It is easy to verify that with this restriction, after translating from $\{0,1\}$ to $\{1,-1\}$ and appropriate normalization, all the linear constraints in the corresponding integer program as in Equation 6.3 are ε-regular. Thus, using the generator from Theorem 1.5 and enumerating over all seeds to the generator, we have the following:

Theorem 6.1. *There exists a deterministic algorithm that, given instance of an ε-dense set covering problem with k constraints over a universe of size n, approximates the number of solutions to within an additive factor of δ in time $n^{\mathrm{poly}(\log k, 1/\delta)}$ as long as $\varepsilon \leq \delta^5/(\log^{8.1} k)(\log(1/\delta))$.*

We now briefly elaborate on approximately counting the number of $\{0,1\}$ contingency tables. The problem of counting $\{0,1\}$-contingency tables is the following. Given, positive integers n, k $n > k$, $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{Z}^n$, $\mathbf{c} = (c_1, \ldots, c_k) \in \mathbb{Z}^k$ we wish to count the number of solutions, $\mathsf{CT}(\mathbf{r}, \mathbf{c})$, to the following integer program whose solutions are matrices $X \in \{0,1\}^{n \times k}$ with row and column sums given by $\mathbf{r}, \mathbf{c}$.

$$\begin{aligned} &\text{Find } X \in \{0,1\}^{n \times k} \\ \text{s.t. } &\sum_j X_{ij} = r_i, 1 \leq i \leq n, \\ &\sum_i X_{ij} = c_j, 1 \leq j \leq k. \end{aligned}$$

Observe that, after translating from $\{0,1\}$ to $\{1,-1\}$ and appropriately normalizing, solutions to the above integer program correspond to points from $\{1,-1\}^{n\times k}$ that lie in an intersection of $2(n+k)$-halfspaces each of which is $(1/\sqrt{k})$-regular (recall that the notion of regularity does not depend on the value of the c_i's or r_j's). Thus, as with dense instances of set cover, we can use Theorem 1.5 to count the number of $\{0,1\}$-contingency tables:

Theorem 6.2. *There exists a deterministic algorithm that on input $\mathbf{r}\in\mathbb{Z}^n$, $\mathbf{c}\in\mathbb{Z}^k$, approximates $\mathsf{CT}(\mathbf{r},\mathbf{c})/2^{nk}$, the fraction of $\{0,1\}$-contingency tables with sums $\mathbf{r},\mathbf{c}$, to within additive error δ, and runs in time $n^{\mathrm{poly}(\log k,1/\delta)}$.*

We remark that using results of Wolff [Wol07], who shows hypercontractivity for various discrete distributions, we can approximately count number of solutions to dense set cover instances and contingency tables over most natural domains.

6.3 Pseudorandom Generators for Polytopes in Gaussian Space

We now prove Theorem 1.6. We use an idea of Ailon and Chazelle [AC06] and the invariance of the Gaussian measure to unitary rotations to obtain PRGs with respect to $\mathcal{N}^n$ for *all* polytopes. Similar ideas were used by Meka and Zuckerman to obtain PRGs for spherical caps (i.e., the case of one hyperplane). In our setting, we must prove that, with respect to a random rotation, *all* of the bounding hyperplanes become regular with high probability. Such a tail bound requires applying hypercontractivity.

Let $H\in\mathbb{R}^{n\times n}$ be the normalized Hadamard matrix with $H_{ij}\in\{1/\sqrt{n},-1/\sqrt{n}\}$. Ailon and Chazelle show that for any $w\in\mathbb{R}^n$, and a random diagonal matrix D with uniformly random $\{1,-1\}$ entries, the vector HDw is regular with high probability. We derandomize their observation using hypercontractivity. For a vector $x\in\mathbb{R}^n$, let $D(x)\in\mathbb{R}^{n\times n}$ be the diagonal matrix with diagonal entries x.

Lemma 6.3. *There exists a constant $C>0$ such that the following holds. For any $w\in\mathbb{R}^n$, $\|w\|=1$, $0<\delta<1$ and any $(C\log(k/\delta))$-wise independent distribution $\mathcal{D}$ over $\{1,-1\}^n$,*

$$\Pr_{x\leftarrow\mathcal{D}}\left[\,\|HD(x)w\|_4^4\geq C\log^2(k/\delta)/n\,\right]\leq\delta/k.$$

Proof. Fix a $w\in\mathbb{R}^n$ and a $C\log(k/\delta)$-wise independent distribution $\mathcal{D}$ for constant C to be chosen later. Let random variable $Z=\|HD(x)w\|_4^4=\sum_i\left(\sum_l H_{il}x_lw_l\right)^4$ for $x\leftarrow\mathcal{D}$. Note that x satisfies $(2,q)$-hypercontractivity for $q\leq C\log(k/\delta)$. Now,

$$\begin{aligned}
\mathbb{E}[Z^2]&=\sum_{i,j}\mathbb{E}\left[\left(\sum_l H_{il}x_lw_l\right)^4\left(\sum_{l'}H_{jl'}x_{l'}w'_l\right)^4\right]\\
&\leq\sum_{ij}\sqrt{\mathbb{E}\left[\left(\sum_l H_{il}x_lw_l\right)^8\right]\cdot\mathbb{E}\left[\left(\sum_l H_{jl}x_lw_l\right)^8\right]}\qquad\text{Cauchy-Schwarz inequality}\\
&\leq\sum_{i,j}8^4\left(\mathbb{E}\left[\left(\sum_l H_{il}x_lw_l\right)^2\right]\right)^2\left(\mathbb{E}\left[\left(\sum_l H_{jl}x_lw_l\right)^2\right]\right)^2\qquad(2,8)\text{-hypercontractivity}\\
&=8^4\sum_{i,j}\frac{1}{n^4}=\frac{c}{n^2}.
\end{aligned}$$

Observe that Z is a degree 4 multilinear polynomial over $x_1, \ldots, x_n$. Therefore, by $(2,q)$-hypercontractivity, Lemma 2.4, applied to the random variable Z, for $q \leq C \log(k/\delta)/4$,

$$\mathbb{E}[|Z|^q] \leq q^{2q} (\mathbb{E}[Z^2])^{q/2} \leq \frac{c^{q/2}\, q^{2q}}{n^q}.$$

Hence, by Markov's inequality, for $\gamma > 0$,

$$\mathsf{Pr}[\,|Z| > \gamma\,] = \mathsf{Pr}[\,|Z|^q > \gamma^q\,] \leq \left(\frac{c^{1/2}\, q^2}{\gamma n}\right)^q.$$

The lemma now follows by taking $q = 2\log(k/\delta)$ and $\gamma = 2\, c^{1/2}\, q^2/n$. □

Let $G : \{0,1\}^r \to \{1,-1\}^n$ be the generator from Theorem 1.5 for $r = O((\log n \log k)/\varepsilon)$. Let $G_1 : \{0,1\}^{r_1} \to \{1,-1\}^n$ generate a $C\log(k/\delta)$-wise independent distribution, for constant C as in Lemma 6.3. Generators G_1 as above with $r_1 = O(\log(k/\delta)\log n)$ are known. Define $G_{\mathcal{N}} : \{0,1\}^{r_1} \times \{0,1\}^r \to \mathbb{R}^n$ as follows:

$$G_{\mathcal{N}}(x,y) = D(G_1(x))HG(y).$$

We claim that $G_{\mathcal{N}}$ δ-fools all polytopes with respect to $\mathcal{N}^n$.

Proof of Theorem 1.6. Recall that $\varepsilon = \Omega(\delta^{5.1}/\log^{8.1} k) > 1/n^{.51}$. The seed length of $G_{\mathcal{N}}$ is $r_1 + r = O(\log n \log k/\varepsilon)$. Fix $W \in \mathbb{R}^{n\times n}$. Observe that $W^T G_{\mathcal{N}}(x,y) = (HD(G_1(x))W)^T G(y)$. Now, from Lemma 6.3 and a union bound it follows that

$$\mathop{\mathsf{Pr}}_{x\in_u\{0,1\}^{r_1}}[\,HD(G_1(x))W \text{ is not } \varepsilon\text{-regular}\,] \leq \delta. \tag{6.4}$$

Further, from the invariance of $\mathcal{N}^n$ with respect to unitary rotations, for any $x \in \{0,1\}^{r_1}$,

$$\mathop{\mathsf{Pr}}_{z\leftarrow\mathcal{N}^n}[(HD(G_1(x))W)^T z \in \mathsf{Rect}(\theta)] = \mathop{\mathsf{Pr}}_{z\leftarrow\mathcal{N}^n}[W^T z \in \mathsf{Rect}(\theta)].$$

Thus, from Theorem 1.5 applied to $\mathcal{N}$, we get that for $HD(G_1(x))W$ ε-regular,

$$\left|\mathop{\mathsf{Pr}}_{y\in_u\{0,1\}^r}[\,(HD(G_1(x))W)^T G(y) \in \mathsf{Rect}(\theta)\,] - \mathop{\mathsf{Pr}}_{z\leftarrow\mathcal{N}^n}[W^T z \in \mathsf{Rect}(\theta)]\right| \leq \delta. \tag{6.5}$$

The theorem now follows from Equations (6.4), (6.5). □

6.4 Pseudorandom Generators for Intersections of Spherical Caps

Theorem 1.7 follows from Theorem 1.6 and the following new invariance principle for polytopes over S^{n-1}:

Lemma 6.4. *For any polytope $\mathcal{K}$ with k faces,*

$$\left|\mathop{\mathsf{Pr}}_{X\in_u S^{n-1}}[X \in \mathcal{K}] - \mathop{\mathsf{Pr}}_{Y\leftarrow\mathcal{N}^n}\left[Y/\sqrt{n} \in \mathcal{K}\right]\right| \leq \frac{C\log n\, \log k}{\sqrt{n}}.$$

The proof uses Nazarov's bound on Gaussian surface area and the following classical large deviation bound for the norm of a random Gaussian vector (for a nice exposition of the bound see [Tao09])

Lemma 6.5. *For $Y \leftarrow \mathcal{N}^n$,*

$$\mathsf{Pr}[\,|\|Y\| - \sqrt{n}| > t\,] \le a\exp(-b\,t^2),$$

where $a, b > 0$ are universal constants.

Proof of Lemma 6.4. Fix a polytope $\mathcal{K}(W,\theta)$. Let $X \in_u S^{n-1}$ and $Y \leftarrow \mathcal{N}^n$. Note that $Y/\|Y\|$ is uniformly distributed over S^{n-1}. Fix $\delta = c/n^{1/2}$ for a constant c to be chosen later. Observe that for $Y \leftarrow \mathcal{N}^n$, and $u \in \mathbb{R}^n$, $\|u\| = 1$, $\langle u, Y\rangle$ is distributed as $\mathcal{N}$. Hence, for any $u \in \mathbb{R}^n, \|u\| = 1$,

$$\mathsf{Pr}[\,|\langle u, Y\rangle| \ge \sqrt{\log(k/\delta)}\,] \le \frac{\delta}{k}.$$

Therefore, by a union bound,

$$\mathsf{Pr}[\,\|W^T Y\|_\infty/\sqrt{n} > \sqrt{\log(k/\delta)}/\sqrt{n}\,] \le \delta.$$

Further, by using Lemma 6.5 and the fact that $Y/\|Y\|$ is uniformly distributed over S^{n-1},

$$\mathsf{Pr}[\,\|W^T X\|_\infty > \sqrt{C\log(k/\delta)}/\sqrt{n}\,] \le 2\delta,$$

for a sufficiently large constant C. From the above two equations, it follows that to prove the theorem we can assume that

$$\|\theta\|_\infty < \sqrt{C\log(k/\delta)/n}.$$

Now, from Lemma 6.5 and the above equation it follows that

$$\mathsf{Pr}[\,|\|Y\| - \sqrt{n}|\,\|\theta\|_\infty \ge \sqrt{C\log(1/\delta)\log(k/\delta)/n}\,] \le \delta. \tag{6.6}$$

Let $\lambda = \sqrt{C\log(1/\delta)\log(k/\delta)/n}$. Then, since $Y/\|Y\| \in_u S^{n-1}$

$$\begin{aligned}
|\,\mathsf{Pr}[X \in \mathcal{K}] - \mathsf{Pr}[Y/\sqrt{n} \in \mathcal{K}]\,| &= |\,\mathsf{Pr}[W^T X \in \mathsf{Rect}(\theta)] - \mathsf{Pr}[W^T Y/\sqrt{n} \in \mathsf{Rect}(\theta)]\,| \\
&= |\,\mathsf{Pr}[W^T Y \in \|Y\|\,\mathsf{Rect}(\theta)] - \mathsf{Pr}[W^T Y \in \sqrt{n}\,\mathsf{Rect}(\theta)]\,| \\
&\le \mathsf{Pr}[\,|\|Y\| - \sqrt{n}|\,\|\theta\|_\infty \ge \lambda\,] \\
&\quad + \mathsf{Pr}[\,W^T Y \in \mathsf{Rect}(\sqrt{n}\theta + \lambda 1_k) \setminus \mathsf{Rect}(\sqrt{n}\theta - \lambda 1_k)\,] \\
&\le \delta + O(\,\lambda\sqrt{\log k}\,). \qquad \text{(Equation 6.6, Lemma 3.4)}
\end{aligned}$$

The lemma now follows by choosing $\delta = c/n^{1/2}$ for a sufficiently large constant c. □

Proof of Theorem 1.7. Define $G_{sp} : \{0,1\}^{r_1} \times \{0,1\}^r \to S^{n-1}$ by $G_{sp}(x,y) = G_{\mathcal{N}}(x,y)/\sqrt{n}$. It follows from Theorem 1.6 and Lemma 6.4 that G_{sp} fools polytopes over S^{n-1} as in the theorem. □

Acknowledgments

Thanks to Fedja Nazarov for helping us compute an integral. We had useful conversations with Carly Klivans, Ryan O'Donnell, Alistair Sinclair, Eric Vigoda, and David Zuckerman.

References

[AC06] Nir Ailon and Bernard Chazelle. *Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform.* In *Proc. 38th ACM Symp. on Theory of Computing (STOC)*, pages 557–563. 2006. doi:10.1145/1132516.1132597.

[Ben90] Vidmantas K. Bentkus. *Smooth approximations of the norm and differentiable functions with bounded support in Banach space l_∞^k.* Lithuanian Mathematical Journal, 30(3):223–230, July 1990. doi:10.1007/BF00970805.

[Ben03] ———. *On the dependence of the Berry-Esseen bound on dimension.* Journal of Statistical Planning and Inference, 113(2):385–402, May 2003. doi:10.1016/S0378-3758(02)00094-0.

[BKS99] Itai Benjamini, Gil Kalai, and Oded Schramm. *Noise sensitivity of Boolean functions and applications to percolation.* Inst. Hautes Études Sci. Publ. Math., 90(1):5–43, 1999. arXiv:math/9811157, doi:10.1007/BF02698830.

[BR07] Matthias Beck and Sinai Robins. *Computing the Continuous Discretely: Integer-point Enumeration in Polyhedra*. Undergraduate Texts in Mathematics. Springer, 1st edition, 2007.

[BV08] Alexander Barvinok and Ellen Veomett. *The computational complexity of convex bodies*. In Jacob E. Goodman, János Pach, and Richard Pollack, eds., *Surveys on Discrete and Computational Geometry: Twenty Years Later*, volume 453 of *Contemporary Mathematics*, pages 117–137. AMS, 2008. arXiv:math/0610325.

[CD03] Mary Cryan and Martin E. Dyer. *A polynomial-time algorithm to approximately count contingency tables when the number of rows is constant.* J. Computer and System Sciences, 67(2):291–310, 2003. doi:10.1016/S0022-0000(03)00014-X.

[DKN09] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. *Bounded independence fools degree-2 threshold functions*, 2009. arXiv:0911.3389.

[Dye03] Martin E. Dyer. *Approximate counting by dynamic programming.* In *Proc. 35th ACM Symp. on Theory of Computing (STOC)*, pages 693–699. 2003. doi:10.1145/780542.780643.

[Fel68] William Feller. *An Introduction to Probability Theory and Its Applications, Volume 1.* Wiley, 3rd edition, 1968.

[Fel71] ———. *An Introduction to Probability Theory and Its Applications, Volume 2.* Wiley, 2nd edition, 1971.

[GOWZ09] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. *Fooling functions of halfspaces under product distributions*, 2009. (Manuscript).

[GW95] Michel X. Goemans and David P. Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.* J. ACM, 42(6):1115–1145, 1995. (Preliminary version in *26th STOC*, 1994). doi:10.1145/227683.227684.

[Hås01] Johan Håstad. *Some optimal inapproximability results.* J. ACM, 48(4):798–859, July 2001. (Preliminary Version in *29th STOC*, 1997). doi:10.1145/502090.502098.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. *Pseudorandomness for network algorithms.* In *Proc. 26th ACM Symp. on Theory of Computing (STOC)*, pages 356–364. 1994. doi:10.1145/195058.195190.

[JS97] Mark Jerrum and Alistair Sinclair. *The Markov chain Monte Carlo method: An approach to approximate counting and integration*. In Dorit S. Hochbaum, ed., *Approximation Algorithms for NP-hard Problems*. PWS Publishing Company, 1997.

[Kal05] Gil Kalai. *Noise sensitivity and chaos in social choice theory*. Technical Report 399, Center for Rationality and Interactive Decision Theory, Hebrew University of Jerusalem, 2005.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. *The influence of variables on Boolean functions (extended abstract).* In *Proc. 29th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 68–80. 1988. doi:10.1109/SFCS.1988.21923.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?* SIAM J. Computing, 37(1):319–357, 2007. (Preliminary version in *45th FOCS*, 2004). eccc:TR05-101, doi:10.1137/S0097539705447372.

[KKMS08] Adam Tauman Kalai, Adam R. Klivans, Yishay Mansour, and Rocco A. Servedio. *Agnostically learning halfspaces.* SIAM J. Computing, 37(6):1777–1805, 2008. (Preliminary version in *46th FOCS*, 2005). doi:10.1137/060649057.

[KOS04] Adam R. Klivans, Ryan O'Donnell, and Rocco A. Servedio. *Learning intersections and thresholds of halfspaces*. J. Computer and System Sciences, 68(4):808–840, 2004. (Preliminary version in *43rd FOCS*, 2002). doi:10.1016/j.jcss.2003.11.002.

[KOS08] ———. *Learning geometric concepts via Gaussian surface area*. In *Proc. 49th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 541–550. 2008. doi:10.1109/FOCS.2008.64.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. *Constant depth circuits, Fourier transform, and learnability*. J. ACM, 40(3):607–620, 1993. (Preliminary version in *30th FOCS*, 1989). doi:10.1145/174130.174138.

[Man94] Yishay Mansour. *Learning Boolean functions via the Fourier transform*. In Vwani P. Roychowdhury, Kai-Yeung Siu, and Alon Orlitsky, eds., *Theoretical Advances in Neural Computation and Learning*, pages 391–424. Kluwer Academic Publishers, 1994.

[MH99] Sanjeev Mahajan and Ramesh Hariharan. *Derandomizing approximation algorithms based on semidefinite programming*. SIAM J. Computing, 28(5):1641–1663, 1999. (Preliminary version in *36th FOCS*, 1995). doi:10.1137/S0097539796309326.

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. *Noise stability of functions with low in.uences invariance and optimality*. In *Proc. 46th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 21–30. 2005. arXiv:math/0503503, doi:10.1109/SFCS.2005.53.

[Mos08] Elchanan Mossel. *Gaussian bounds for noise correlation of functions and tight analysis of long codes*. In *Proc. 49th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 156–165. 2008. arXiv:math/0703683, doi:10.1109/FOCS.2008.44.

[MZ09] Raghu Meka and David Zuckerman. *Pseudorandom generators for polynomial threshold functions*, 2009. arXiv:0910.4122.

[Naz03] Fedor Nazarov. *On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure*. In *Geometric Aspects of Functional Analysis (Israel Seminar 2001–2002)*, volume 1807/2003 of *Lecture Notes in Mathematics*, pages 169–187. Springer, 2003. doi:10.1007/b10415.

[NN93] Joseph Naor and Moni Naor. *Small-bias probability spaces: Efficient constructions and applications*. SIAM J. Computing, 22(4):838–856, August 1993. (Preliminary Version in *22nd STOC*, 1990). doi:10.1137/0222053.

[O'D04] Ryan O'Donnell. *Hardness amplification within NP*. J. Computer and System Sciences, 69(1):68–94, 2004. (Preliminary version in *34th STOC*, 2002). doi:10.1016/j.jcss.2004.01.001.

[O'D08] ———. *Some topics in analysis of Boolean functions*. In *Proc. 40th ACM Symp. on Theory of Computing (STOC)*, pages 569–578. 2008. eccc:TR08-055, doi:10.1145/1374376.1374458.

[Per04] Yuval Peres. *Noise stability of weighted majority*, 2004. arXiv:math/0412377.

[Pin94] Iosif Pinelis. *Extremal probabilistic problems and hotellings T^2 test under a symmetry condition*. Ann. Statist., 22(1):357–368, 1994. doi:10.1214/aos/1176325373.

[PR89] Vygantas Paulauskas and Alfredas Račkauskas. *Approximation Theory in the Central Limit Theorem: Exact Results in Banach Spaces*. Kluwer Academic Publishers, 1989. (Translated from Russian).

[Shi00] Yaoyun Shi. *Lower bounds of quantum black-box complexity and degree of approximating polynomials by influence of Boolean variables*. Inf. Process. Lett., 75(1-2):79–83, 2000. arXiv:quant-ph/9904107, doi:10.1016/S0020-0190(00)00069-7.

[Tao09] Terry Tao. *Talagrand's concentration inequality*, 2009. (Post in Blog "What's new").

[Wol07] Pawel Wolff. *Hypercontractivity of simple random variables*. Studia Math, 180(3):219–236, 2007. doi:10.4064/sm180-3-3.

[Wol08] Ronald de Wolf. *A brief introduction to Fourier analysis on the Boolean cube*. Theory of Computing, Graduate Surveys, 1:1–20, 2008. doi:10.4086/toc.gs.2008.001.

[Zie95] Günter M. Ziegler. *Lectures on polytopes*, volume 152 of *Graduate texts in Mathematics*. Springer, 1995.